

Testing and modeling users



DILBERT reprinted by permission of United Feature Syndicate.

The aims

- Describe how to do user testing
- Discuss the difference between user testing, usability testing and research experiments
- Discuss the role of user testing in usability testing
- Discuss how to design simple experiments
- Describe GOMS, the keystroke model, Fitt's law and discuss when these techniques are useful
- Describe how to do a keystroke level analysis

Experiments, user testing & usability testing

- Experiments tests hypotheses to discover new knowledge by investigating the relationship between two or more things – i.e variables
- User testing is applied experimentation in which developers check that the system being developed is usable by the intended user population for their tasks
- Usability testing uses a combination of techniques including user testing & satisfaction questionnaires

User testing is not research

- **User testing**
 - Aim: improve products
 - Few participants
 - Results inform design
 - Not perfectly replicable
 - Controlled conditions
 - Procedure planned
 - Results reported to developers
- **Research**
 - Aim: discover knowledge
 - Many participants
 - Results validated statistically
 - Replicable
 - Strongly controlled conditions
 - Experimental design
 - Scientific paper reports the results to the community

User testing

- Goals & questions focus on how well users perform tasks with the product
- Comparison of products or prototypes common
- Major part of usability testing
- Focus is on time to complete a task & number & type of errors
- Informed by video and interaction logging
- User satisfaction questionnaires provide data about users' opinions

Testing conditions

- Usability lab or other controlled space
- Major emphasis on:
 - Selecting representative users
 - Developing representative tasks
- 5-10 users typically selected
- Tasks usually last no more than 30 minutes
- The test conditions should be the same for each participant
- Informed consent form explains ethical issues

Type of data

(Wilson & Wixon, 97)

- Time to complete a task
- Time to complete a task after a specified time away from a product
- Number and type of errors per task
- Number of errors per unit of time
- Number of navigations to online help or manuals
- Number of users making a particular error
- Number of users completing task successfully

Usability engineering specifications

- Current level of performance
- Minimum acceptable level of performance
- Target level of performance

How many participants is enough for user testing ?

- The number is largely a practical issue
- Depends on:
 - Schedule for testing
 - Availability of participants
 - Cost of running tests
- Typical 5- 10 participants
- Some experts argue that testing should continue until no new insights are gained

Experiments

- Predict the relationship between two or more variables
- Independent variable is manipulated by the researcher
- Dependent variable depends on the independent variable
- Typical experimental designs have one or two independent variables
- Hypothesis: “Will the time to read a screen of text be different if 12-point Helvetica font is used instead of 12-point Times New Roman ?
- **IMPORTANT:** ordering effects, learning curve

Experimental designs

- Different participants – single group of participants is allocated randomly to the experimental conditions
- Some participants – all participants appear in both conditions
- Matched participants – participants are matched in pairs, e.g., based on expertise, gender

Advantages & disadvantages

Design	Advantages	Disadvantages
Different	No order effects	Many subjects & individual differences a problem
Same	Few individuals, no individual differences	Counter-balancing needed because of ordering effects
Matched	Same as different participants but individual differences reduced	Cannot be sure of perfect matching on all differences

Predictive models

- Provide ways of evaluating designs or products without directly involving users
- Psychological models of users are used to test designs
- Less expensive than user testing
- Usefulness limited to systems with predictable tasks – e.g. Telephone answering systems

GOMS

(Card et al, 1983)

- Goals – the state the user wants to achieve e.g.
Find a website
- Operators – the cognitive processes & physical actions performed to attain these goals, e.g.
Decide which search engine to use
- Methods – the procedures for accomplishing the goals e.g drag the mouse over field, type in keywords, press go button
- Selection rules – determine which method to select when there is more than one available
- Example: Delete word in sentence (Word)

Keystroke model

- Quantitative predictions
- K pressing single key or button 0.35 (avg)
 - Skilled typist (55 wpm) – 0.22
 - Average typist (40 wpm) – 0.28
 - User unfamiliar with the keyboard – 1.20
- Pointing with a mouse or other device to a target on a display – 1.10
- Clicking the mouse – 0.20
- Mentally prepare to do something – 1.35

Benefits and limitations of GOMS

- Comparative analyses for different interfaces
- Project Ernestine (Gray et al. 1993)
 - Empirical data collected
 - Compared to GOMS analysis
- Limitations
 - Highly limited scope : small set of highly routine data-entry type tasks
 - No account of errors
 - Multitasking, interruptions

Fitt's Law (1954)

- The law predicts that the time to point at an object using a device is a function of the distance from the target object & the object's size
- The further away & the smaller the object, the longer the time to locate the point
- Useful for evaluating systems for which the time to locate an object is important such as handheld devices such as mobile phones

Fitt's law

- $T = k \log_2 (D/S + 0.5)$, $k \sim 100$ msec
- T = time to move the hand to a target
- D = distance between hand and target
- S = size of target
- Nokia : 12-key mobile phone keypad

Key points

- User testing is a central part of usability engineering
- Testing is done in controlled conditions
- User testing is an adapted form of experimentation
- Experiments aim to test hypotheses by manipulating certain variables while keeping others constant
- The experimenter controls the independent variable(s) but not the dependent variable(s)
- There are three types of experimental design: different participants, same participants & matched participants
- GOMS, Keystroke model and Fitt's law predict expert error-free performance
- Predictive models are used to evaluate systems with predictable tasks such as telephones